Sigfrido Gomez

Geog 579

23 July 2017

Lab 4: Kriging

**QUESTION 1**: Explain what lag size and number of lag mean? How will they affect the semivarigram. (2 points)

A large lag size will mask short-range autocorrelation and if it’s too small there may be many empty bins. These bins will be too small to get representative averages. The number of lags specify how many lags of the variogram to calculate. I used the average nearest neighbor for the lag size. In choosing the number of lag, the rule of thumb is to multiply the lag size by the number of lags, which should be about half the largest distance among all points. However, depending on whether the points are clustered or dispersed it is advisable to increase the lag size.

**QUESTION 2**: Choose two best semivariogram models from the 4 commonly-used models: Circular, Spherical, Exponential and Gaussian. Explain why you think they are the best for describing the sample semivariogram. (2 points)

Seeing that our sample size is dispersed I believe that using the Gaussian model puts more emphasis on the weights to closer samples and makes more use to build interpolation. The other model to choose would be the spherical since it does not vary much from the sill of the Gaussian model. The exponential spreads the weights more evenly that gives a general interpolation though most of the weights.

**QUESTION 3**: Explain how “Maximum Neighbors”, “Minimum neighbors” and “Sector type” affect the searing neighborhood (2 points).

Sector type is used to alter the type of searching neighborhood using 1,4,4 with an of set of 45 degrees or 8 sectors. Using the crosshairs, we see the weights of the neighbors in a sector and how they will be used to predict an unmeasured location. The minimum number of features includes the minimum features needed for interpolation and if not met, no prediction can be made. However, widespread practice is to change the search of the neighborhood to include the minimum. Maximum neighbors play an important role when different sectors are used and the maxi for kriging is 200. It’s common to have the max set with the number of sectors.

**QUESTION 4**: Describe what you see in “Predicted” and “Error” scatterplots. Analyze the cross validation result by using those two scatter plots as well as the summary shown under the scatter plot (2 points).

Both the prediction and error plot are the same except the measured values are subtracted from the predicted values. In the predicted value we are seeing that the blue line does not line up with the grey so we are seeing that the kriging model here is poor because it should be closer to 1:1 gray line in the predicted plot.

**QUESTION 5**: Examine the interpolation variance maps carefully. Describe how the interpolation variance varies over space (from a sample point to locations without sampling, and overall variation across the map) on a single map and explain why it varies in that fashion. (2 points)

Areas with higher density of features have a low predicted standard error. The areas that have lower and no features at all are filled with high prediction errors. These areas are filled in with in between features and for this particular sample the higher prediction areas look like filled in squares. The larger the distance in-between features the higher the prediction.

**QUESTION 6**: Compare the results from the RMSE for difference interpolations and judge which interpolation is better. Compare and discuss this judgment with what you concluded in Question 2. (2 points)

RMSE is high at around 137 since I used many of the default options. I believe that my first attempt was wrong and an exponential approach may be a better approach. Looking at the values produce, the exponential results are slightly lower and reduce the RMSE.

**QUESTION 7**: Print the interpolation variance image and compare it with the one produced with the default neighborhood sizes. Run RMSE and compare the results with those from the default neighborhood size. Discuss the differences (both in variance maps and in the rmse and me results). (2 points)

The RSME is higher in the exponential testing that it is with Gaussian. Also, the maps are very similar in how much the density in-between features is filled with error. The major difference is that the exponential map has slightly higher standard error than the Gaussian method. Experimenting with the circular method garnered comparable results as Gaussian. The spaces in-between was not as filled with error on the edges that touched areas that had little to no samples points. The RSME was still very high for both the exponential and circular methods, 139.80 and 136.98 respectively. As with testing different models, I had over looked that the shape of the curve should be as closely aligned with the averaged points. So, the circular method would be the best model for this data set even as it reduces the prediction error. With the lowest RSME and changes in the neighborhood sizes do help in the reduction of the RSME and close in the gap for error.